

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-337815

(43)Date of publication of application : 06.12.1994

(51)Int.Cl. G06F 12/08
G06F 12/12
G06F 13/12

(21)Application number : 05-127280

(71)Applicant : HITACHI LTD

(22)Date of filing : 28.05.1993

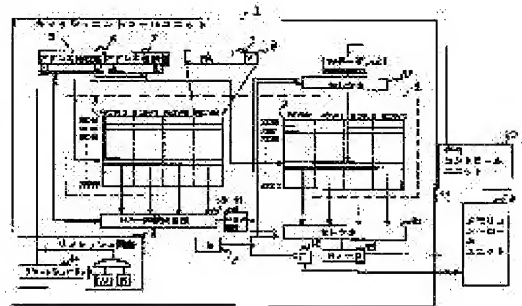
(72)Inventor : KASHIYAMA YUKI

(54) DATA PROCESSOR

(57)Abstract:

PURPOSE: To obtain a high hit ratio with inexpensive constitution in a space which is physically restricted by constituting a cache memory with DRAM and providing a flag storage means storing a flag showing the validity of data in the cache memory.

CONSTITUTION: The cache memory 4 in a cache control unit 1 is constituted by DRAM. When a memory read request is given, real address data 8 and a valid flag 9 are read out of the address array buffer 3 of the cache memory 4. The address data 8 and the valid flag 9 are inputted to a matching detection circuit 10, and they are compared with address data of an address comparison part 5. When the valid flag 9 is '1' (data is valid) and matched address data exists as the result of matching, ROW information 11 which is cache-hit is inputted to a selector 16, and data which is read out of the corresponding ROW position of a cache buffer 2 is selected.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] In the data processor equipped with the cache controller which controls cache memory and this cache memory between main storage and a data processing unit A flag storage means to memorize the flag which constitutes said cache memory from a DRAM, and shows the effectiveness of the data in cache memory in said cache controller for every access unit of data, A flag setting means to set up the applicable flag of said flag storage means effectively at the time of access of data, The data processor characterized by establishing an output means to output as a signal which shows read-out for the flag which corresponds from said flag storage means at the time of access of data, and shows the validity of access data, and an invalid for the flag.

[Claim 2] The data processor according to claim 1 characterized by reading the data for access from main storage only when said flag shows an invalid.

[Claim 3] The data processor according to claim 1 characterized by preparing further the refresh circuit refreshed for said cache memory only in the state of standby of a data processing unit in a cache controller.

[Claim 4] The data processor of claim 1 characterized by constituting said data processing unit, cache memory, and a cache controller from a single chip LSI thru/or 3 publications either.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Industrial Application] This invention relates to the data processor equipped with the cache

controller which controls cache memory and this cache memory between main storage and a data processing unit.

[0002]

[Description of the Prior Art] In the former, in order to raise the transfer rate of the data (an instruction and data) to a data processing unit from main storage If cache memory is prepared between main storage and a data processing unit, main storage is accessed and data are read It writes also in cache memory at the same time it transmits the read-out data to a data processing unit. The data processor with cache memory which transmitted applicable data to the data processing unit from cache memory is known without accessing main storage, when the data for access exist in cache memory at the time of the next access.

[0003] By the way, the need of securing considerable storage capacity in such a data processor in order to gather the hit ratio over cache memory is **. For example, if it is in on-line processing, it is desirable to secure the storage capacity of 2MB or more.

[0004] On the other hand, in recently, the demand to single-chip-izing and downsizing of a data processing unit and cache memory increases, and to aim at a deployment of cache memory in the space restricted physically is desired.

[0005]

[Problem(s) to be Solved by the Invention] However, if it is in a data processor with the conventional cache memory, SRAM is used as cache memory as indicated by the publication-number No. 296541 official report. For this reason, although the access time is short compared with DRAM, a degree of integration is small, since it is expensive, it is the space restricted physically and it is becoming difficult to secure the storage capacity to need. Consequently, there is a problem that a hit ratio is high in the space restricted physically, and an effective data processing unit cannot be built in respect of cost.

[0006] The purpose of this invention is to offer the data processor which is the space restricted physically and can acquire a high hit ratio with a cheap configuration.

[0007]

[Means for Solving the Problem] In order to attain said purpose, this invention notes that the access time of the latest DRAM is becoming a high speed extremely like 30ns. A flag storage means to memorize the flag which constitutes cache memory from a DRAM and shows the effectiveness of the data in cache memory in said cache controller for every access unit of data, A flag setting means to set up the applicable flag of said flag storage means effectively at the time of access of data, It is characterized by establishing an output means to output as a signal which shows read-out for the flag which corresponds from said flag storage means at the time of access

of data, and shows the validity of access data, and an invalid for the flag.

[0008] Furthermore, it is characterized by preparing further the refresh circuit refreshed for said cache memory only in the state of standby of a data processing unit in a cache controller.

[0009]

[Function] According to the above-mentioned means, the flag of a flag storage means shows the "invalid" according to the initial state first. However, if data are written in the cache address of arbitration, in parallel to this write-in actuation, it will be set as "effective" by the flag setting means.

[0010] If the cache address corresponding to the flag which is [this] "effective" is again accessed within the refresh time amount specified in the property table of DRAM (read-out or actuation of writing), it will be reactivated the same with having refreshed in the refresh circuit, and data will be held by this repetition. At this time, a flag is also again set as an "effective" condition.

[0011] However, if there is no access even if it passes over the refresh time amount specified in the property table of DRAM, the data of the cache address corresponding to the flag which is "effective" will disappear automatically. In connection with this, a flag is set as an "invalid" condition.

[0012] On the other hand, if cache memory is accessed, the flag corresponding to the access unit will be outputted. Since it means that the data for access had existed in cache memory when this flag shows the "effective" condition, the read-out demand to main storage is not generated. However, when the "invalid" condition is shown, while the flag itself is used as a read-out demand to main storage, and the data for access are read from main storage and transmitted to a data processing unit, it is stored also to cache memory.

[0013] Namely, data will be maintained if access to DRAM is performed within the refresh time amount specified in the property table in this invention. If are not carried out within refresh time amount, and it uses that data carry out automatic disappearance and access from a data processing unit is performed within the refresh time amount specified in the property table of DRAM The effectiveness of the data in cache memory is guaranteed, effectiveness is not guaranteed, but the description is to have made it the configuration of reading the data for access from main storage. Therefore, when a data processing unit is always working, the refresh circuit of dedication is not needed. Moreover, a means to cancel old data is not needed on conditions in which the cache address which is in the "invalid" condition always exists more than the specified quantity, either.

[0014] In addition, only when a data processing unit will be in a standby condition, the refresh circuit of dedication is added.

[0015]

[Example] Hereafter, it explains to a detail based on the example illustrating this invention.

[0016] Drawing 1 is the block diagram showing one example of this invention, and shows the configuration detailed about the cache control unit which is the important section of this invention here.

[0017] The cache memory 4 which consists of the cache buffer 2 the cache control unit 1 of this example remembers data to be, and the address array buffer 3 which memorizes the access address data of the data memorized by this cache buffer 2 is formed.

[0018] This cache buffer 2 and the address array buffer 3 consist of DRAMs. And the cache buffer 2 and the address array buffer 3 are set as the storage capacity of 4 M bytes by which 4 lows (ROW) and each low data width of face were constituted [the line writing direction] for the 512K column and the direction of a train from 2 bytes. In this case, each column of a cache buffer 2 and the address array buffer 3 and a low correspond to 1 to 1, respectively, and one low in each column serves as an access unit of data.

[0019] Moreover, the access address register 5 which memorizes the access address data from a data processing unit (not shown) is formed in the cache control unit 1.

[0020] Since the access address data from a data processing unit are 32 bits, this access address register 5 consists of 32-bit registers, from bit [the 0th byte of] 1 to bit [the 1st byte of] 2 is used as an address comparator 6 among those, and from bit [the 1st byte of] 3 of that low order to bit [the 3rd byte of] 4 is used as the address retrieval section 7. That is, the address retrieval section 7 is used as column address data of a cache buffer 2 and the address array buffer 3, and the address comparator 6 of the high order is used for the comparison with the address data read from 4 lows of an applicable column.

[0021] Here, 11 bits of high orders of the access address data when accessing main storage, i.e., the address data of the same configuration as the address comparator 6, are written in ROW0-ROW3 of each column of the address array buffer 3 as real address data (RA) 8, and the BARIDDO flag 9 which shows that the data of the storage location corresponding to 1 bit of the low order in a cache buffer 2 are still more effective is written in them.

[0022] So, when the address data of the address comparator 6 are compared with the address data read from 4 lows of the applicable column of the address array buffer 3 and there are congruous address data, the real address coincidence detector 10 which outputs the ROW information 11 showing the ROW location and the in cache information 12 (it is "1" with data) which shows that the data for access existed in cache memory 4 is formed.

[0023] Furthermore, the refresh circuit 13 for refreshing cache memory 4 is formed. This refresh circuit 13 is equipped with the refresh address generating circuit 14 which carries out refresh

address data output, and only when the system-wait signal W which shows that a data processing unit is waiting is "1", or only when refresh request signal R is "1", it is constituted so that refresh address data may be outputted for every fixed time amount.

[0024] Only when a data processing unit is waiting, he is trying to output refresh address data here for preventing that cache access occurs frequently while a data processing unit is working, and compete with this, and the engine performance falls.

[0025] On the other hand, based on the in cache information 12 and the ROW information 11, the data of ROW which carried out the cache hit are chosen, and the selector 16 outputted as read-out data 15 and the selector which chooses ROW which carried out the cache hit are prepared in the cache buffer 2 side.

[0026] Furthermore, the in cache information 12 is reversed and the knot in cache information Nin is inputted into the memory control unit 19 as a read-out demand signal to main storage (or memory of the high order hierarchy of this cache memory). And if the data for access are read by the read-out demand signal, the read-out data will be inputted into a register 21 through the instruction control unit 20, and will be written in the ROW location of the column with which a cache buffer 2 corresponds through a selector 17. At this time, 11 bits of high orders of the access address are written in the ROW location of the column with which the BARIDDO flag 9 of "1" is specified in the address retrieval section 7 of access address data as real address data 8.

[0027] Actuation concerning the above configuration is explained below.

[0028] First, when a memory read-out demand is outputted from a data processing unit, the in cache information 12 is set to "0" in the initial state according to which data do not exist in cache memory 4. Therefore, the knot in cache information Nin is set to "1."

[0029] This knot in cache information Nin is inputted into the memory control unit 19 as a read-out demand signal to main storage (or memory of the high order hierarchy of this cache memory).

[0030] In parallel to this, 11 bits of high orders of the access address data inputted into the access address register 5 are written in the ROW location of the column with which the real address data 8 and the BARIDDO flag 9 of "1" are specified in the address retrieval section 7 of the access address.

[0031] On the other hand, the data for [which was read from main storage by the read-out demand signal] access are inputted into a register 21 through the instruction control unit 20, and it is written in the ROW location of the column with which a cache buffer 2 corresponds through a selector 17.

[0032] Next, when a memory read-out demand is outputted from a data processing unit, in the condition that data exist in cache memory 4, the column position of the address array buffer 3 is

specified by the address retrieval section 7, and 4 sets of real address data 8 and the BARIDDO flag 9 are read from the column.

[0033] These 4 sets of real address data 8 and the BARIDDO flag 9 are inputted into the coincidence detector 10, and are compared with the address data of the address comparator 5. If there are address data which the BARIDDO flag 9 is "1" and are in agreement as a result of this comparison, since it will mean that the access address concerned was accessed within the refresh time amount of DRAM, the in cache information 12 having been set to "1" with the BARIDDO flag 9 of "1", and the knot in cache information Nin having been further set to "0", and having carried out the cache hit at the memory control unit 19 and the instruction control unit 20 is notified. Since it means that the column which carried out the cache hit was accessed within the refresh time amount of DRAM at this time, the contents of the original address data are activated without refreshing specially.

[0034] The ROW information 11 which carried out the cache hit is inputted into a selector 16, the data read from the ROW location where a cache buffer 2 corresponds are chosen as coincidence, and it is inputted into it as read-out data 15 at the instruction control unit 20. Also in this cache buffer 2, since it means that the column which carried out the cache hit was accessed within the refresh time amount of DRAM, the contents of the original data are activated, without refreshing specially.

[0035] Next, when a write request occurs from a data processing unit, it writes in the access address register 5 and address data are inputted, like the case where it is read-out, the column position of the address array buffer 3 is specified by the address retrieval section 7, and 4 sets of real address data 8 and the BARIDDO flag 9 are read from the column.

[0036] These 4 sets of real address data 8 and the BARIDDO flag 9 are inputted into the coincidence detector 10, and are compared with the address data of the address comparator 5. If there are data which the BARIDDO flag 9 is "1" and are in agreement as a result of this comparison, since it will mean that the access address concerned was accessed within the refresh time amount of DRAM, the in cache information 12 having been set to "1" with the BARIDDO flag 9 of "1", and the knot in cache information Nin having been further set to "0", and having carried out the cache hit at the memory control unit 19 and the instruction control unit 20 is notified. And it writes in and data 21 are written in the ROW location where the ROW

information 11 which carried out the cache hit was inputted into the selector 17, and was inputted through the instruction control unit 20 and where a cache buffer 2 corresponds by the selector 17.

[0037] In this case, since it means that the column in which the address array buffer 3 carried out the cache hit was accessed within the refresh time amount of DRAM, the contents of the original

address data are activated, without refreshing specially.

[0038] However, since it means that the column position concerned is the condition that it is accessed exceeding the refresh time amount of DRAM, and data cannot be guaranteed, or an initial state when the BARIDDO flag 9 is "0" as a result of the comparison in the coincidence detector 10, the address data of the address comparator 5 are written in the empty ROW location of the column concerned as real address data 8, and the BARIDDO flag 9 of "1" is further written in it at the low order. On the other hand, it writes in the ROW location where a cache buffer 2 corresponds, and data 21 are written in.

[0039] On the other hand, if a data processing unit will be in a standby condition and the system-wait signal W is set to "1", refresh address data will be generated in the fixed time interval of under the refresh time amount of DRAM from the refresh address generating circuit 14, and sequential refresh of a cache buffer 2 and the address array buffer 3 will be carried out per column.

[0040] By this refresh actuation, the data and address data in a cache buffer 2 and the address array buffer 3 are held so that it may not disappear in the standby condition of a data processing unit.

[0041] Refresh request signal R occurs and this is the same also at the time.

[0042] By the way, in a multiprocessor system, although coincidence elimination of the data in cache memory may be compulsorily carried out for coincidence control of data, since it is realizable by setting the BARIDDO flags 9 to "0" all at once, in the cache memory 4 of this example, there is an advantage that improvement in the speed can be attained.

[0043] Thus, in this example, only when the BARIDDO flag 9 which constitutes cache memory 4 from a DRAM, and shows the effectiveness of the data in cache memory 4 in the cache control unit 1 is made to memorize for every access unit of data, and the applicable BARIDDO flag 9 is effectively set up for every access of data and the applicable BARIDDO flag 9 shows the invalid at the time of access of data, it constitutes so that data may be updated.

[0044] Namely, data will be maintained if access to DRAM is performed within the refresh time amount specified in the property table. If are not carried out within refresh time amount, and it uses that data carry out automatic disappearance and access from a data processing unit is performed within the refresh time amount specified in the property table of DRAM The effectiveness of the data in cache memory 4 is guaranteed, effectiveness is not guaranteed, but it is made the configuration of reading the data for access from main storage.

[0045] Therefore, when a data processing unit is always working, the refresh circuit of dedication is not needed. Moreover, on conditions in which the cache address which is in the "invalid"

condition always exists more than the specified quantity, there is an advantage of not needing a means to cancel old data, either.

[0046] Moreover, only when a data processing unit is waiting, by being made to perform refresh actuation, contention with cache access of a data processing unit is lost, and there is an advantage that it can prevent that the engine performance falls.

[0047] It is the space restricted especially physically, and there is an advantage that a high hit ratio can be acquired with a cheap configuration. For example, when the cache memory which consisted of 256 K bytes of SRAM is transposed to the cache memory of 4 M bytes of DRAM, storage capacity increases 16 times and the frequency which does not carry out a cache hit falls remarkably. Consequently, the overhead at the time of the block transfer of data also decreases sharply. And cost also falls.

[0048] Therefore, if the cache memory control unit 1 and a data processing unit are constituted as a single chip LSI, it is cheap and the microprocessor of high performance can be realized.

[0049] In addition, although the flag which shows that data exist in cache memory, and the BARIDDO flag which shows whether data are effective are shared in this example, it is good even if separate. What is necessary is to notify that data are effective, only when it is made separate, and both flags are "1."

[0050] Moreover, you may constitute so that the address retrieval section 7 may be divided into two steps, a column address and the class address, and the address array buffer 3 may be searched.

[0051]

[Effect of the Invention] As explained above, in this invention, cache memory is constituted from a DRAM. And the flag which shows the effectiveness of the data in cache memory in a cache control unit is made to memorize for every access unit of data. Since it constituted so that data might be updated only when an applicable flag was effectively set up for every access of data and the applicable flag showed the invalid at the time of access of data When a data processing unit is always working, in the refresh circuit of dedication becoming unnecessary, on conditions in which the cache address which is in the "invalid" condition always exists more than the specified quantity, it is effective in not needing a means to cancel old data, either.

[0052] Moreover, only when a data processing unit is waiting, by being made to perform refresh actuation, contention with cache access of a data processing unit is lost, and it is effective in the ability to prevent that the engine performance falls.

[0053] It is the space restricted especially physically, and a high hit ratio can be acquired with a cheap configuration, and it is effective in the ability to decrease sharply the overhead at the time of the block transfer of data.

[0054] Moreover, if a cache memory control unit and a data processing unit are constituted as a single chip LSI, it is effective in it being cheap and being able to realize the microprocessor of high performance.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

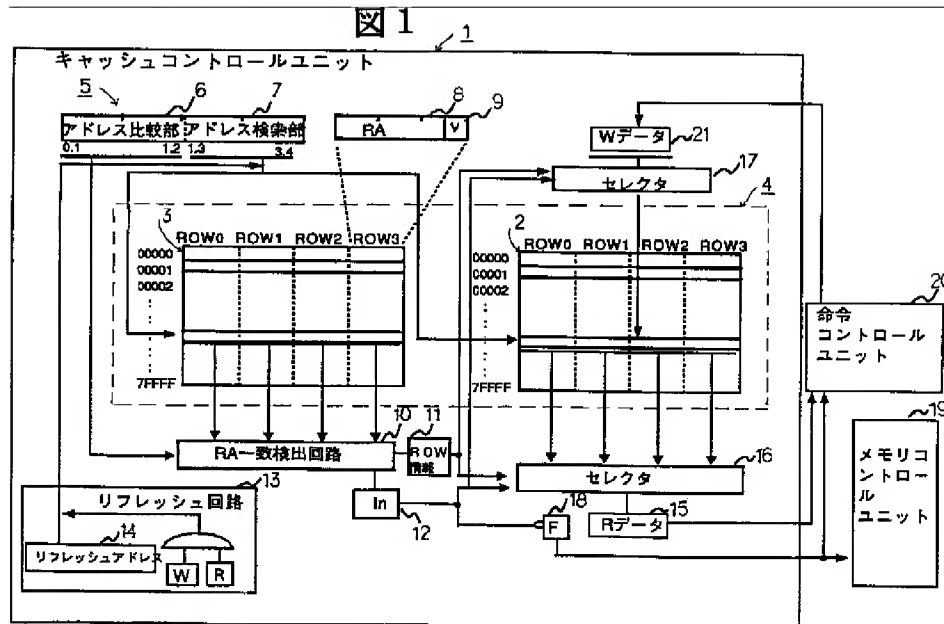
[Drawing 1] It is the block block diagram showing one example of this invention.

[Description of Notations]

1 [-- A BARIDDO flag, 12 / -- In cache information, 13 / -- Refresh circuit.] -- A cache control unit, 2 -- A cache buffer, 3 -- An address array buffer, 9

DRAWINGS

[Drawing 1]



(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平6-337815

(43)公開日 平成6年(1994)12月6日

(51)Int.Cl. ⁵	識別記号	序内整理番号	P I	技術表示箇所
G 0 6 F	12/08	3 1 0 Z	7608-5B	
	12/12	A	7608-5B	
	13/12	3 3 0 T	8133-5B	

審査請求 未請求 請求項の数4 O L (全 6 頁)

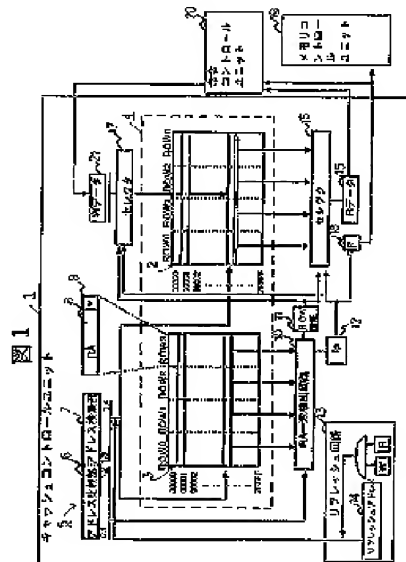
(21)出願番号	特願平5-127230	(71)出願人	000005108 株式会社日立製作所 東京都千代田区神田駿河台四丁目6番地
(22)出願日	平成5年(1993)5月28日	(72)発明者	柏山 由紀 神奈川県秦野市堀山下1番地 株式会社日立製作所汎用コンピュータ事業部内
		(74)代理人	弁理士 秋田 収喜

(54)【発明の名称】 データ処理装置

(57)【要約】

【目的】 物理的に限られた空間で、かつ安価な構成で高いヒット率を得ることができるようにする。

【構成】 キャッシュメモリをD R A Mで構成し、かつキャッシュコントロールユニット内に、キャッシュメモリ内のデータの有効性を示すフラグをデータのアクセス単位毎に記憶させ、データのアクセス毎に該当フラグを有効に設定するようにし、かつデータのアクセス時に該当フラグが無効を示している時のみデータを更新する。



【特許請求の範囲】

【請求項1】 主記憶装置とデータ処理ユニットとの間にキャッシュメモリおよび該キャッシュメモリを制御するキャッシュコントローラを備えたデータ処理装置において、

前記キャッシュメモリをDRAMで構成し、かつ前記キャッシュコントローラ内に、キャッシュメモリ内のデータの有効性を示すフラグをデータのアクセス単位毎に記憶するフラグ記憶手段と、データのアクセス時に前記フラグ記憶手段の該当フラグを有効に設定するフラグ設定手段と、データのアクセス時に前記フラグ記憶手段から該当するフラグを読み出し、そのフラグをアクセスデータの有効、無効を示す信号として出力する出力手段とを設けたことを特徴とするデータ処理装置。

【請求項2】 前記フラグが無効を示す時のみ主記憶装置からアクセス対象のデータを読み出すことを特徴とする請求項1記載のデータ処理装置。

【請求項3】 データ処理ユニットの待機状態でのみ前記キャッシュメモリをリフレッシュするリフレッシュ回路をキャッシュコントローラ内にさらに設けたことを特徴とする請求項1記載のデータ処理装置。

【請求項4】 前記データ処理ユニット、キャッシュメモリおよびキャッシュコントローラをシングルチップLSIで構成したことを特徴とする請求項1ないし3記載のいずれかのデータ処理装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、主記憶装置とデータ処理ユニットとの間にキャッシュメモリおよび該キャッシュメモリを制御するキャッシュコントローラを備えたデータ処理装置に関するものである。

【0002】

【従来の技術】 従来において、主記憶装置からデータ処理ユニットに対するデータ（命令またはデータ）の転送速度を向上させるために、主記憶装置とデータ処理ユニットとの間にキャッシュメモリを設け、主記憶装置をアクセスしてデータを読み出したならば、その読出しデータをデータ処理ユニットに転送すると同時にキャッシュメモリへも書き込んでおき、次のアクセス時にアクセス対象のデータがキャッシュメモリに存在していた場合は、主記憶装置をアクセスすることなく、キャッシュメモリから該当データをデータ処理ユニットに転送するようにしたキャッシュメモリ付のデータ処理装置が知られている。

【0003】 ところで、このようなデータ処理装置においては、キャッシュメモリに対するヒット率を上げるためには、相当の記憶容量を確保しておく必要がある。例えば、オンライン処理にあっては2MB以上の記憶容量を確保しておくことが望ましい。

【0004】 一方、最近においては、データ処理ユニッ

トおよびキャッシュメモリのシングルチップ化やダウンサイジングに対する要求が高まり、物理的に限られた空間でキャッシュメモリの有効利用を図ることが望まれている。

【0005】

【発明が解決しようとする課題】 しかし、従来のキャッシュメモリ付のデータ処理装置にあっては、特開平296541号公報に開示されているように、キャッシュメモリとしてSRAMを使用している。このため、DRAMに比べてアクセス時間は短い、集積度が小さく、高価であるために物理的に限られた空間で、必要とする記憶容量を確保することが困難になってきている。この結果、物理的に限られた空間でヒット率が高く、コストの点でも有効なデータ処理ユニットを構築することができないという問題がある。

【0006】 本発明の目的は、物理的に限られた空間で、かつ安価な構成で高いヒット率を得ることができるデータ処理装置を提供することにある。

【0007】

【課題を解決するための手段】 前記目的を達成するために本発明は、最近のDRAMのアクセス時間が30nsというように極めて高速になってきていることに注目し、キャッシュメモリをDRAMで構成し、かつ前記キャッシュコントローラ内に、キャッシュメモリ内のデータの有効性を示すフラグをデータのアクセス単位毎に記憶するフラグ記憶手段と、データのアクセス時に前記フラグ記憶手段の該当フラグを有効に設定するフラグ設定手段と、データのアクセス時に前記フラグ記憶手段から該当するフラグを読み出し、そのフラグをアクセスデータの有効、無効を示す信号として出力する出力手段とを設けたことを特徴とする。

【0008】 さらに、データ処理ユニットの待機状態でのみ前記キャッシュメモリをリフレッシュするリフレッシュ回路をキャッシュコントローラ内にさらに設けたことを特徴とする。

【0009】

【作用】 上記手段によれば、まず、フラグ記憶手段のフラグは初期状態では「無効」を示している。しかし、任意のキャッシュアドレスにデータが書き込まれると、この書き込み動作と並行してフラグ設定手段によって「有効」に設定される。

【0010】 この「有効」状態となっているフラグに対応するキャッシュアドレスがDRAMの特性表で規定されたりリフレッシュ時間以内に再びアクセス（読出し、または書き込みの動作）されると、リフレッシュ回路でリフレッシュしたのと同様に再活性化され、この繰返しによってデータが保持される。このとき、フラグも「有効」状態に再度設定される。

【0011】 しかし、DRAMの特性表で規定されたりリフレッシュ時間を過ぎても、アクセスがなければ、「有

効」状態となっているフラグに対応するキャッシュアドレスのデータは自動的に消滅する。これに伴って、フラグは「無効」状態に設定される。

【0012】一方、キャッシュメモリをアクセスすると、そのアクセス単位に対応したフラグが出力される。もし、このフラグが「有効」状態を示していた場合は、アクセス対象のデータがキャッシュメモリ内に存在していたことになるので、主記憶装置への読出し要求は発生しない。しかし、「無効」状態を示していた時には、フラグ自身が主記憶装置への読出し要求として使用され、主記憶装置からアクセス対象データが読み出され、データ処理ユニットに転送されるとともに、キャッシュメモリに対しても格納される。

【0013】すなわち、本発明においては、DRAMに対するアクセスが特性表で規定されたリフレッシュ時間以内に行われればデータが維持され、リフレッシュ時間以内に行われなければデータは自動消滅することを利用し、データ処理ユニットからのアクセスがDRAMの特性表で規定されたリフレッシュ時間以内に行われれば、キャッシュメモリ内のデータの有効性を保証し、そうでなければ有効性を保証せず、主記憶装置からアクセス対象データを読み出すという構成にしたことに特徴がある。従って、データ処理ユニットが常時稼働中である場合は、専用のリフレッシュ回路を必要としない。また、「無効」状態となっているキャッシュアドレスが常に所定量以上存在するような条件では、古いデータを破棄する手段も必要としない。

【0014】なお、データ処理ユニットが待機状態になる場合のみ、専用のリフレッシュ回路が付加される。

【0015】

【実施例】以下、本発明を図示する実施例に基づいて詳細に説明する。

【0016】図1は本発明の一実施例を示すブロック図であり、ここでは本発明の要部であるキャッシュコントロールユニットについて詳細な構成を示している。

【0017】この実施例のキャッシュコントロールユニット1は、データを記憶するキャッシュバッファ2と、このキャッシュバッファ2に記憶されたデータのアクセスアドレスデータを記憶するアドレスレイバッファ3とから成るキャッシュメモリ4が設けられている。

【0018】このキャッシュバッファ2およびアドレスレイバッファ3は、DRAMで構成されている。そして、キャッシュバッファ2およびアドレスレイバッファ3は、行方向が512Kカラム、列方向が4 ROW、各 ROW のデータ幅が2バイトで構成された4Mバイトの記憶容量に設定されている。この場合、キャッシュバッファ2およびアドレスレイバッファ3の各カラムおよび ROW はそれぞれ1対1に対応し、各カラムの中の1つの ROW がデータのアクセス単位となる。

【0019】また、キャッシュコントロールユニット1

には、データ処理ユニット（図示せず）からのアクセスアドレスデータを記憶するアクセスアドレスレジスタ5が設けられている。

【0020】このアクセスアドレスレジスタ5は、データ処理ユニットからのアクセスアドレスデータが32ビットであることから、32ビットのレジスタで構成され、そのうち0バイト目のビット1から1バイト目のビット2までがアドレス比較部6として用いられ、その下位の1バイト目のビット3から3バイト目のビット4までがアドレス検索部7として用いられる。すなわち、アドレス検索部7は、キャッシュバッファ2およびアドレスレイバッファ3のカラムアドレスデータとして用いられ、その上位のアドレス比較部6は該当カラムの4 ROW から読み出されるアドレスデータとの比較のために用いられる。

【0021】ここで、アドレスレイバッファ3の各カラムのROW0～ROW3には、主記憶装置をアクセスした時のアクセスアドレスデータの上位11ビット、すなわちアドレス比較部6と同一構成のアドレスデータがリアルアドレスデータ（RA）8として書き込まれ、さらにその下位1ビットに、キャッシュバッファ2の対応する記憶位置のデータが有効であることを示すバリッドフラグ9が書き込まれるようになっている。

【0022】そこで、アドレス比較部6のアドレスデータとアドレスレイバッファ3の該当カラムの4 ROW から読み出されるアドレスデータとを比較し、一致したアドレスデータがあった時には、そのROW位置を表すROW情報11と、キャッシュメモリ4内にアクセス対象データが存在したことを示すインキャッシュ情報12（データ有りで“1”）を出力するリアルアドレス一致検出回路10が設けられている。

【0023】さらに、キャッシュメモリ4をリフレッシュするためのリフレッシュ回路13が設けられている。このリフレッシュ回路13はリフレッシュアドレスデータ出力するリフレッシュアドレス発生回路14を備え、データ処理ユニットが待機中であることを示すシステムウェイト信号Wが“1”の時のみ、またはリフレッシュリクエスト信号Rが“1”の時のみ一定時間毎にリフレッシュアドレスデータを出力するように構成されている。

【0024】ここで、データ処理ユニットが待機中の時のみリフレッシュアドレスデータを出力するようにしているのは、データ処理ユニットが稼働中の時は頻繁にキャッシュアクセスが発生し、これと競合して性能が低下することを防止するためである。

【0025】一方、キャッシュバッファ2側には、インキャッシュ情報12とROW情報11に基づき、キャッシュヒットしたROWのデータを選択し、読出しデータ15として出力するセクタ16と、キャッシュヒットしたROWを選択するセクタが設けられている。

10

20

30

40

50

【0026】さらに、インキャッシュ情報12を反転し、ノットインキャッシュ情報Ninはメモリコントロールユニット19に主記憶装置（または該キャッシュメモリの上位階層のメモリ）への読出し要求信号として入力される。そして、読出し要求信号によってアクセス対象のデータが読み出されると、その読出しデータは命令コントロールユニット20を介してレジスタ21に入力され、セレクト17を通じてキャッシュバッファ2の該当するカラムのROW位置に書き込まれる。この時、アクセスアドレスの上位11ビットがリアルアドレスデータ8として、また“1”のバリッドフラグ9がアクセスアドレスデータのアドレス検索部7で指定されるカラムのROW位置に書き込まれる。

【0027】以上の構成に係る動作を以下説明する。

【0028】まず、メモリ読出し要求がデータ処理ユニットから出力された時、キャッシュメモリ4にデータが存在していない初期状態では、インキャッシュ情報12は“0”となる。従って、ノットインキャッシュ情報Ninは“1”となる。

【0029】このノットインキャッシュ情報Ninはメモリコントロールユニット19に主記憶装置（または該キャッシュメモリの上位階層のメモリ）への読出し要求信号として入力される。

【0030】これと並行して、アクセスアドレスレジスタ5に入力されたアクセスアドレスデータの上位11ビットがリアルアドレスデータ8と、“1”のバリッドフラグ9がアクセスアドレスのアドレス検索部7で指定されるカラムのROW位置に書き込まれる。

【0031】一方、読出し要求信号によって主記憶装置から読み出されたアクセス対象のデータは命令コントロールユニット20を介してレジスタ21に入力され、セレクト17を通じてキャッシュバッファ2の該当するカラムのROW位置に書き込まれる。

【0032】次に、メモリ読出し要求がデータ処理ユニットから出力された時、キャッシュメモリ4にデータが存在している状態では、アドレス検索部7によってアドレスアレイバッファ3のカラム位置が指定され、そのカラムから4組のリアルアドレスデータ8とバリッドフラグ9が読み出される。

【0033】これら4組のリアルアドレスデータ8とバリッドフラグ9は一致検出回路10に入力され、アドレス比較部5のアドレスデータと比較される。この比較の結果、バリッドフラグ9が“1”で、かつ一致するアドレスデータがあれば、当該アクセスアドレスはDRAMのリフレッシュ時間以内にアクセスされたことになるので、“1”のバリッドフラグ9によってインキャッシュ情報12が“1”となり、さらにノットインキャッシュ情報Ninが“0”となってメモリコントロールユニット19および命令コントロールユニット20にキャッシュヒットしたことが通知される。この時、キャッシュヒ

ットしたカラムはDRAMのリフレッシュ時間以内にアクセスされたことになるので、特別にリフレッシュすることなく元のアドレスデータの内容に活性化される。

【0034】同時に、キャッシュヒットしたROW情報11がセレクト16に入力され、キャッシュバッファ2の対応するROW位置から読み出されたデータが選択され、読出しデータ15として命令コントロールユニット20に入力される。このキャッシュバッファ2においても、キャッシュヒットしたカラムはDRAMのリフレッシュ時間以内にアクセスされたことになるので、特別にリフレッシュすることなく元のデータの内容に活性化される。

【0035】次にデータ処理ユニットから書き込み要求が発生し、アクセスアドレスレジスタ5に書き込みアドレスデータが入力された場合、読出しの場合と同様に、アドレス検索部7によってアドレスアレイバッファ3のカラム位置が指定され、そのカラムから4組のリアルアドレスデータ8とバリッドフラグ9が読み出される。

【0036】これら4組のリアルアドレスデータ8とバリッドフラグ9は一致検出回路10に入力され、アドレス比較部5のアドレスデータと比較される。この比較の結果、バリッドフラグ9が“1”で、かつ一致するデータがあれば、当該アクセスアドレスはDRAMのリフレッシュ時間以内にアクセスされたことになるので、“1”のバリッドフラグ9によってインキャッシュ情報12が“1”となり、さらにノットインキャッシュ情報Ninが“0”となってメモリコントロールユニット19および命令コントロールユニット20にキャッシュヒットしたことが通知される。そして、キャッシュヒットしたROW情報11がセレクト17に入力され、命令コントロールユニット20を通じて入力された書き込みデータ21がセレクト17によってキャッシュバッファ2の対応するROW位置に書き込まれる。

【0037】この場合、アドレスアレイバッファ3のキャッシュヒットしたカラムはDRAMのリフレッシュ時間以内にアクセスされたことになるので、特別にリフレッシュすることなく元のアドレスデータの内容に活性化される。

【0038】しかし、一致検出回路10における比較の結果、バリッドフラグ9が“0”であった場合、当該カラム位置はDRAMのリフレッシュ時間を過ぎてアクセスされてデータが保証できない状態、または初期状態であることを意味するので、当該カラムの空きROW位置に、アドレス比較部5のアドレスデータをリアルアドレスデータ8として書き込み、さらにその下位に“1”のバリッドフラグ9を書き込む。一方、キャッシュバッファ2の該当するROW位置には書き込みデータ21を書き込む。

【0039】一方、データ処理ユニットが待機状態になり、システムウエイト信号Wが“1”になると、リフレ

ッシュアドレス発生回路14からDRAMのリフレッシュ時間未満の一定時間間隔でリフレッシュアドレスデータが発生され、キャッシュバッファ2およびアドレスアレイバッファ3はカラム単位で順次リフレッシュされる。

【0040】このリフレッシュ動作によってキャッシュバッファ2およびアドレスアレイバッファ3内のデータおよびアドレスデータは、データ処理ユニットの待機状態においても消滅しないように保持される。

【0041】これは、リフレッシュリクエスト信号Rが10発生して時も同様である。

【0042】ところで、マルチプロセッサシステム等では、データの一致制御のために、キャッシュメモリ内のデータを強制的に同時消去することがあるが、この実施例のキャッシュメモリ4においては、バリッドフラグ9を一斉に“0”にすることによって実現することができるので、高速化が図れるという利点がある。

【0043】このように本実施例においては、キャッシュメモリ4をDRAMで構成し、かつキャッシュコントロールユニット1内に、キャッシュメモリ4内のデータの有効性を示すバリッドフラグ9をデータのアクセス単位毎に記憶させ、データのアクセス毎に該当バリッドフラグ9を有効に設定するようにし、かつデータのアクセス時に該当バリッドフラグ9が無効を示している時のみデータを更新するように構成したものである。

【0044】すなわち、DRAMに対するアクセスが特性表で規定されたリフレッシュ時間以内に行われればデータが維持され、リフレッシュ時間以内に行われなければデータは自動消滅することを利用し、データ処理ユニットからのアクセスがDRAMの特性表で規定されたリフレッシュ時間以内に行われれば、キャッシュメモリ4内のデータの有効性を保証し、そうでなければ有効性を保証せず、主記憶装置からアクセス対象データを読み出すという構成にしたものである。

【0045】従って、データ処理ユニットが常時稼動中である場合は、専用のリフレッシュ回路を必要としない。また、「無効」状態となっているキャッシュアドレスが常に所定量以上存在するような条件では、古いデータを破棄する手段も必要としないという利点がある。

【0046】また、データ処理ユニットが待機中の時のみリフレッシュ動作を実行するようにすることにより、データ処理ユニットのキャッシュアクセスとの競合がなくなり、性能が低下することを防止できるという利点がある。

【0047】特に、物理的に限られた空間で、かつ安価な構成で高いヒット率を得ることができるという利点がある。例えば、256KバイトのSRAMで構成されたキャッシュメモリを4MバイトのDRAMのキャッシュメモリに置き換えた場合、記憶容量は16倍になり、キャッシュヒットしない頻度が著しく低下する。この結

果、データのブロック転送時のオーバーヘッドも大幅に減少する。しかも、コストも低下する。

【0048】したがって、キャッシュメモリコントロールユニット1とデータ処理ユニットをシングルチップLSIとして構成すれば、安価で高性能のマイクロプロセッサを実現することができる。

【0049】なお、この実施例では、キャッシュメモリ内にデータが存在することを示すフラグと、データが有効であるかどうかを示すバリッドフラグとを共用しているが、別々にしてもよい。別々にした場合は、両方のフラグが“1”のときのみデータが有効であることを通知するようにすればよい。

【0050】また、アドレス検索部7をカラムアドレスとクラスアドレスの2段に分けてアドレスアレイバッファ3を検索するように構成してもよい。

【0051】

【発明の効果】以上説明したように本発明においては、キャッシュメモリをDRAMで構成し、かつキャッシュコントロールユニット内に、キャッシュメモリ内のデータの有効性を示すフラグをデータのアクセス単位毎に記憶させ、データのアクセス毎に該当フラグを有効に設定するようにし、かつデータのアクセス時に該当フラグが無効を示している時のみデータを更新するように構成したので、データ処理ユニットが常時稼動中である場合は、専用のリフレッシュ回路が不要になるうえ、「無効」状態となっているキャッシュアドレスが常に所定量以上存在するような条件では、古いデータを破棄する手段も必要としないという効果がある。

【0052】また、データ処理ユニットが待機中の時のみリフレッシュ動作を実行するようにすることにより、データ処理ユニットのキャッシュアクセスとの競合がなくなり、性能が低下することを防止できるという効果がある。

【0053】特に、物理的に限られた空間で、かつ安価な構成で高いヒット率を得ことができ、データのブロック転送時のオーバーヘッドを大幅に減少させることができるという効果がある。

【0054】また、キャッシュメモリコントロールユニットとデータ処理ユニットをシングルチップLSIとして構成すれば、安価で高性能のマイクロプロセッサを実現することができるという効果がある。

【図面の簡単な説明】

【図1】本発明の一実施例を示すブロック構成図である。

【符号の説明】

1…キャッシュコントロールユニット、2…キャッシュバッファ、3…アドレスアレイバッファ、9…バリッドフラグ、12…インキャッシュ情報、13…リフレッシュ回路。

【図1】

